On the assessment of merging processes for the improvement of the spatial resolution of multispectral SPOT XS images

Thierry Ranchin, Lucien Wald, Marc Mangolini, Claude Penicand

ON THE ASSESSMENT OF MERGING PROCESSES FOR THE IMPROVEMENT OF THE SPATIAL RESOLUTION OF MULTISPECTRAL SPOT XS IMAGES

Thierry RANCHIN (1), Lucien WALD (1), Marc MANGOLINI (2), Claude PENICAND (3)

(1) Groupe Télédétection & Modélisation, Centre d'Energétique, Ecole des Mines de Paris, BP 207, 06904 Sophia Antipolis cedex, France. Tel. 33-93 95 74 53. Fax. 33-93 95 75 35.
(2) Aérospatiale, BP 99, 06332 Cannes la Bocca cedex, France.
(3) Cellule d'Etudes en Géographie Numérique, Délégation Générale à l'Armement, 16 bis av. Prieur de la Côte-d'Or, 92114 Arcueil cedex, France. Tel. 33-16-1-42 31 99 37.

ABSTRACT

Methods have been proposed to produce SPOT multispectral images with enhanced spatial resolution (10 m) owing to the presence of a panchromatic band. Some of them aim at preserving the multispectral content when increasing the spatial resolution. This communication defines the properties of such enhanced multispectral images. It then proposes both a formal approach and some criteria to provide a quantitative assessment of the spectral quality of these products. Five sets of criteria qualify the performances of a method to synthesize the radiometry in a single spectral band as well as the multispectral information when increasing the spatial resolution. The influence of the type of landscape present in the scene upon the assessment of the quality is underlined, as well as its dependence with scale. The whole approach is illustrated by the case of several SPOT images and the ARSIS concept.

1. INTRODUCTION

Many methods have been proposed for the merging of high spectral and high spatial resolutions data in order to produce multispectral images having the highest spatial resolution available within the data set. Some aim at providing a synthetic image close to reality when enhancing the spatial resolution. However none of them proposes any assessment of the quality of the resulting synthetic images but Mangolini *et al.* (1992, 1993) and Munechika *et al.* (1993). The quality of the resulting synthetic images is usually assessed by visual inspection. It is a powerful means to locate major drawbacks, or to detect trends or regional artifacts. Therefore it is a mandatory step to accomplish. However, beyond the visual inspection, mathematical criteria are needed.

We propose both an approach and some criteria to provide a quantitative assessment of the quality. In doing this, we assume that the main demand of the user regards the quality of the transformation of the multispectral content when increasing the spatial resolution. The better the simulation of the spectral content at the enhanced resolution, the more accurate the classification for mapping purposes.

Of particular interest here is the method developed for SPOT case by Ranchin *et al.* (1993, 1994) following the ARSIS concept (Mangolini *et al.* 1992, 1993). This method has been applied to several SPOT images. The results obtained for the city of Barcelona (Spain, 11/09/90) and the surroundings of Strasbourg (France, 03/08/90) are used to illustrate the purpose of this communication. The type of landscape present within the image used to assess the quality of a synthesizing method has an influence upon the results. Whatever the method, the more predictable the change in signal with the scale, the better the quality of the final product. Urban areas (Barcelona) and small agricultural lots including villages (Strasbourg) have been selected for illustration because they are certainly the most difficult types of landscape to deal with. They often point out the qualities and drawbacks of algorithms because of the high variability of information in space and spectral band, induced by the diversity of features both in size and nature.

2. A FORMAL APPROACH FOR QUALITY ASSESSMENT OF THE RESULTING SYNTHETIC IMAGES

Let denote the images of lowest resolution by L_l, and the images of highest resolutions by H_h. The subscripts l and h denote the resolution with which the image L or H is considered, respectively low and high resolution. The subscript s (for small) is to be used later: it denotes a resolution which is twice lower than l. The images L_l and H_l are superimposable. Both L_l and H_h have been obtained by a sensor. The merging method aims at constructing synthetic images L^*_h.

These synthetic images must have the three following properties. Firstly, any synthetic image L^*_h, once degraded to its original resolution l should be as identical as possible to the original image L_l. For example, in the case of SPOT, the synthetic image are called XS^*_{10}. Once resampled to 20 m, this image should be as close as possible to the original XS image. Observing this property means that the method takes into account the differences in atmospheric effects affecting the images of lowest and highest resolutions which have not been acquired within the same spectral bands as well as the effects induced by time-lag in image acquisition. Secondly, any synthetic image L^*_h should be as identical as possible to the image L_h the corresponding sensor would observe with the highest resolution h. Thirdly, the multispectral set of synthetic images L^*_h should be as identical as possible to the multispectral set of images L_h the corresponding sensor would observe with the highest resolution h. In the case of SPOT, this set of synthetic images is the triplet ($XS1^*_{10}$, $XS2^*_{10}$, $XS3^*_{10}$). The assessment of the quality of the resulting spatially enhanced spectral images is now equivalent to the verification of these properties.

Checking the first property: *any synthetic image L^*_h, once degraded to its original resolution l should be as identical as possible to the original image L_l.* To achieve this, the synthetic image L^*_h is spatially degraded to an approximate solution L'_l of L_l. If the first property is true, then L'_l is very close to L_l. The difference between both images is computed on a pixel basis. This difference image should be visually analyzed, once converted into bytes. It should be visually compared to the original image in order to detect trends of error, if any, possibly related to the type of landscape. Then some statistical quantities are to be used to quantitatively express the discrepancies between both images. These quantities are similar to the first and second sets of criteria, described later. Actually the ARSIS concept is inherently built to satisfy this first property. Therefore and to avoid duplication of tables, emphasis is put on the second and third properties.

Checking the second property: *any synthetic image L^*_h should be as identical as possible to the image L_h the corresponding sensor would observe with the highest resolution h.* Reading the second and third properties makes feel the difficulties in checking them. Actually mention is made to L_h, an image that would be sensed if the sensor has a better resolution. This image is of course not available, otherwise all the above-cited methods would not have been developed. The difficulty is partly overcome by the following approach:

- The available images H_h and L_l are degraded to respectively H_l and L_s. For SPOT case, the P and

XS images are degraded to P_{20} (20 m resolution) and XS_{40} (40 m resolution). The images H_l and L_s are very close to what the corresponding sensor would have measured with a degraded resolution, as discussed previously.

- Then the synthesizing method under assessment is applied to H_l and L_s. It provides a synthetic image L^*_l (XS^*_{20} in the SPOT case),
- which is compared to the image-truth L_l (*XS* in the SPOT case) by the means of some criteria described below. The numerical comparison should be made preferably in physical units and also in relative values. Thus different tests made over different scenes may be compared.
- This comparison provides an assessment of the quality of L^*_l,
- Finally it is assumed that this quality is fairly similar to that of the synthesized high-resolution image L^*_h. This point will be discussed later.

To assess the quality of L^*_l, a difference is computed between L_l and L^*_l in a similar way than for the first property. After visual inspection, the difference image is reduced to a few statistical parameters which summarize it. There are a large number of candidate parameters. We have computed many for several tens of cases. We have retained some which definitions are well-known to engineers and researchers <u>and</u> which clearly characterize the qualities and drawbacks of a method. They should be the ones to be used for quality assessment by end-users. Mathematicians proposing a new method should use them as reference but also others which are helpful for a better understanding of the characteristics of the method.

Two sets of criteria are proposed to quantitatively summarize the performance of a method in synthesizing an image in one spectral band. The first set of quantities provides a global view on the discrepancies between the original image L_l and the synthetic one L^*_l. It contains:

- the bias as well as its value relative to the mean value of the original image. Recall that the bias is the mean difference and is also equal to the difference between the means of the original image and of the synthetic one. Ideally, the bias should be null,
- the difference in variances (variance of the original image minus variance of the synthetic one) as well as its value relative to the variance of the original image. This difference expresses the quantity of information added or lost during the enhancement of the spatial resolution. For a method providing too much innovations (in the sense of information theory), *i.e.* "inventing" too much information, the difference will be negative because the variance of the synthetic image will be larger than the original variance. In the opposite case, the difference will be positive. Ideally, the variance difference should be null,
- the correlation coefficient between the original and synthetic images. It shows the similarity in small size structures between the original and synthetic images. It should be as close as possible to 1,
- the standard-deviation of the difference image, as well as its value relative to the mean of the original image. It globally indicates the level of error at any pixel. Ideally, it should be null.

The error at pixel level may be more detailed. Let us compute at each pixel the absolute relative error. It is equal to the absolute value of the difference between the original and synthetic values, divided by the original value. Then the histogram of these relative errors is computed. It can be assimilated to the probability density function. Therefore, we can compute the probability of having at a pixel a relative error (in absolute value) lower than a given threshold. This probability denotes the error made at pixel level, and hence indicates the capability of a method to synthesize the small size structures. The closer to 100 % the probability for a given error threshold, the better the synthesis. The ideal value is a probability of 100 % for a null relative error. Here for computer accuracy reasons, the lowest threshold "no relative error or null error" is set to 0.001 %.

Checking the third property: *the multispectral set of synthetic images* L^*_h *should be as identical as possible to the multispectral set of images* L_h *the corresponding sensor would observe with the highest resolution h.* Visual inspection may be made through color compositing of, for example, the first three principal components of the set of images. Both color composites should agree visually. Practically we recommend the following approach. For each spectral band, the L_l and L^*_l images are juxtaposed into a single computer file. The principal components analysis as well as the color coding are performed on this set of files. The projected L_l and L^*_l images are then extracted from these projected files and the color composites are displayed, simultaneously or alternatively, onto the screen. This approach guarantees that the color composites are comparable. Of course, if only three spectral bands are available like in the SPOT case, there is no need to perform a principal components analysis. The advantage of this visual assessment is that it does show trend in errors, if any, possibly related to landscape features. The drawback of it is that it is a subjective assessment and also that this assessment may be limited either by physiological factors, e.g. color contrast perception by humans, or by technical factors e.g. when large number of spectral bands are present. In this case, and if the landscape offers a large variety of objects, the color re-coding of the first three principal components reduces dramatically the differences between L_l and L^*_l images, particularly if these differences are random, *i.e.* not related to a peculiar landscape feature or to a spectral band.

A quantitative assessment can be made using the following three sets of criteria which quantify the performances of a method to synthesize the spectral signatures during the change in spatial resolution. The third set (numbered after the two sets for the second property) deals with the information correlation between the different spectral images taken two by two. This dependence can be expressed by the correlation coefficients, the ideal values being given by the set of original images L_l. As an example, for the case of SPOT, the correlation coefficient between P_{20} and $XS1^*_{20}$ is computed and compared to the correlation coefficient for P_{20} and $XS1_{20}$. This is done for every couple. The fourth group of criteria partly quantifies the synthesis of the actual multispectral *n*-uplets by a method, where *n*-uplet means the vector composed by each of the *n* spectral bands at a pixel. It comprises the number of different *n*-uplets (*i.e.* the number of spectra) observed in the original L_l and in the synthesized L^*_l sets of images, as well as the difference between these numbers. A positive difference means that the synthesized images do not present enough *n*-uplets; a negative difference means too much spectral innovations.

The previous quantities do not guarantee that the synthesized *n*-uplets are the same than in the original image L_l. This last set of criteria assesses the performances in synthesizing the actual *n*-uplets. It deals with the most frequent *n*-uplets, because they are predominant in multispectral classification. For a given threshold in frequency, only are dealt with the *n*-uplets having a frequency (relative number of pixels) greater than this threshold. The threshold is set to 0.01%, 0.05%, 0.1% et 0.5% successively. The greater the threshold, the lower the number of *n*-uplets, but the greater the number of pixels exhibiting one of these *n*-uplets. For each of the *n*-uplets, is computed the difference between the original frequency and the one observed in the synthesized images. These differences are summarized by the following quantities:

- the number of actual *n*-uplets, the number of coincident *n*-uplets in the synthesized images, as well as the difference between these numbers, expressed also relative to the actual number of *n*-uplets,
- the number of concerned pixels, in absolute value as well as relative to the total number of pixels,

- the difference between the above number of pixels for original and synthesized images, also expressed in absolute value relative to the actual number.

3. APPLICATION TO SOME SPOT SUB-SCENES

The P and XS images are degraded to a resolution of respectively 20 and 40 m. Then images are synthesized at 20 m resolution and compared to the original XS images. Table 1 provides the mean and standard-deviation of the radiances for each spectral band as well as the calibration coefficients.

		Mean	Standard-dev.	Calibration
Barcelona	XS1	58	12	1.2181
	XS2	48	15	1.22545
	XS3	55	9	1.29753
	P	53	15	1.39198
Strasbourg	XS1	61	8	0.99862
	XS2	48	14	1.08383
	XS3	92	14	0.95215

Table 1. Mean radiances, standard-deviations and calibration coefficients (in $W.m^{-2}.sr^{-1}.mm^{-1}$).

Tables 2 and 3 provide a global view on the discrepancies between XS and XS^{*}_{20} for each spectral band (checking the second property). The bias is null. The synthetic images are very close to the ideal values but lack of innovation. The very high correlation coefficient denotes the similarity between structures.

	Barcelona			Strasbourg		
	XS1	XS2	XS3	XS1	XS2	XS3
Bias (ideal value: 0) relative to the mean XS value	0.00 0 %	0.00 0 %	0.00 0 %	0.00 0 %	0.00 0 %	0.00 0 %
Actual variance - estimate (ideal value: 0) relative to the actual variance	7 5 %	7 3 %	8 9 %	4 6 %	9 5 %	8 4 %
Correlation coefficient between XS and estimate (ideal value: 1)	0.99	0.99	0.95	0.98	0.99	0.98
Standard-deviation of the differences (ideal value: 0) relative to the mean XS value	1.9 3 %	1.9 4 %	2.7 5 %	1.5 2 %	2.0 4 %	3.1 3 %

Table 2. Statistics on the differences between the original and synthesized images (radiance unit).

The standard-deviation of the differences is weak. Error at pixel level can be more detailed through the second set of criteria (Table 3). Almost all pixels exhibit a relative error less than or equal to 10 %. There is a very high percentage of pixels exhibiting null relative errors: more than 25 % for XS1 and XS2.

The third property deals with the multispectral character of the data set and is verified through the third to fifth sets of criteria. Table 4 shows that the correlations between the spectral bands XS1, XS2, XS3 and P are preserved through the changing of scale.

		0.001	**1**	**2**	**5**	**10**	**20**
Barcelona	XS1	27	28	63	92	99	100
	XS2	26	26	48	86	99	100
	XS3	15	16	43	76	95	100
Strasbourg	XS1	34	34	76	95	100	100
	XS2	25	25	40	81	97	100
	XS3	19	26	54	97	98	100

Table 3. Probability (in per cent) for having in a pixel a relative error less or equal to the thresholds noted in the first row. The ideal value is 100 as early as the first threshold 0.001 %. The relative errors are in absolute value and in per cent.

	Barcelona		Strasbourg	
	original	synthetic	original	synthetic
P - XS1	0.97	0.97	0.97	0.97
P- XS2	0.97	0.97	0.97	0.97
P- XS3	0.35	0.35	-0.40	-0.41
XS1 - XS2	0.97	0.97	0.97	0.97
XS1 - XS3	0.34	0.34	-0.35	-0.36
XS2 - XS3	0.33	0.33	-0.45	-0.46

Table 4. Correlation coefficient between the spectral bands for the original and the synthesized images. The ideal values are those for the original image.

Table 5 reports on some spectral characteristics of the scene, namely the total number of pixels of the image, the number of spectral triplets and the average number of pixels per triplet (ratio of total number of pixels to the number of triplets). The spectral homogeneity is defined as the inverse of the number of triplets and expressed in per cent. It characterizes the spectral diversity of a scene: the greater this parameter, the less the diversity. For a scene exhibiting an unique spectral object (*i.e.* only one triplet), this spectral homogeneity would be 100 %. The Barcelona scene has a value of 0.002 %, which demonstrates its spectral diversity. The performance in synthesizing the multispectral information is partly provided in Table 6, which presents the difference between the actual number of triplets and the number found in the synthesized images. This difference is small for both cases.

	number of pixels	number of triplets	average number of pixels per triplet	spectral homogeneity (in %)
Barcelona	262 144	45 618	5.7	0.002
Strasbourg	262 144	25 561	10.3	0.004

Table 5. Some spectral characteristics of the scenes.

Actually, most of the missing or superfluous triplets have a low frequency, *i.e.* each of them is carried by a few pixels. This is demonstrated in Table 7 which exhibits the performances in synthesizing the most frequent actual triplets. Each triplet under consideration has a frequency of at least 0.01 %. For Barcelona, it corresponds to 26 pixels; these 1,549 triplets have a cumulated frequency of 22 %, that is that 57,096 pixels among 262,144 (total number) carry one of these triplets. Hence synthesizing them is of primary importance for classification purposes. All these triplets as well as their frequencies are retrieved: difference in number of pixels with original is very low.

	Barcelona		Strasbourg	
	original	ARSIS	original	ARSIS
number of triplets	45 618	42 593	25 561	24 765
difference with original (ideal : 0) (in %)	—	3 025 7 %	—	796 3 %

Table 6. Performance in synthesizing the multispectral information. Difference between the actual number of triplets (XS1. XS2. XS3) and the estimates.

	Barcelona		Strasbourg	
	original	ARSIS	original	ARSIS
number of predominant triplets difference with original (ideal : 0) (in %)	1 549 —	1 549 0 0 %	2 349 —	2 349 0 0 %
minimum frequency for each triplet (in pixels) and relative to the total number of pixels	26 0.01 %	26 0.01 %	26 0.01 %	26 0.01 %
number of pixels (relative to the total number of pixels) difference with original (ideal : 0) (en %)	57 096 (22 %) —	57 688 — - 592 - 1 %	130 161 (50 %) —	127 149 — 3 012 2 %

Table 7. Performance in synthesizing the multispectral information. Difference between the actual frequency of a triplet (XS1. XS2. XS3) and its estimate. Only the most frequent triplets are taken into account.

4. EXTRAPOLATION OF QUALITY ASSESSMENT TO THE HIGHEST RESOLUTION

The verification of the second and third properties of the synthetic images has been made on degraded images: we have synthesized multispectral images at a resolution of 20 m. Such an approach palliates the lack of 'truth' images. How can the assessment of quality of the synthetic images be made at the highest resolution (e.g. 10 m in the SPOT case) from that made at the lowest resolution ? In other words, how can one extrapolate the quality assessment made at the lowest resolution to the highest one ? Intuitively, one thinks that except for objects having a size much larger than the resolution, the error should increase with the resolution, since the complexity of a scene increases with the resolution. That is, one may expect the error made at the highest resolution to be greater than that at the lowest resolution. However several recent works have demonstrated the influence of the resolution on the quantification of parameters extracted from satellite imagery (see e.g. Welch *et al.* 1989; Kong and Vidal-Madjar 1988; Woodcock and Strahler 1987; Raffy 1993). All these studies demonstrate that the quality of the assessment of a parameter is an unpredictable function of the resolution.

It follows that we cannot predict the quality of the synthetic images at the highest resolution from the assessments made with synthetic images at the lowest resolution. To illustrate this discussion, we have assessed the quality of a SPOT image synthesized at 40 m, starting from a P image degraded to 40 m and a XS image degraded to 80 m. The scene is the Barcelona one. The results are presented in Tables 8 and 9, under the heading '40 m'. In these Tables are reported the results obtained for 20 m, output from Tables 2 and 3. One can see that for all parameters, the values displayed for 20 m are better than for 40 m. Hence the method provides better estimates in synthesizing images at 20 m resolution than at 40 m. Such comparisons were made for the different methods and for several scenes, comprising mostly urban areas. It has been found in each case that the quality was best at 20 m, and also that the ranking of a method relative to the others was the same at 20 m and 40 m. Though our conclusions were always the same, it does not make a proof that estimates should be better at 10 m than at 20 m. However, we can reasonably assume that the quality of the synthetic images at the highest resolution (e. g. 10 m) is close to that at the lowest resolution (e. g. 20 m).

	20 m	40 m
Bias (ideal value: 0) relative to the mean XS value	0.00 0 %	0.00 0 %
Actual variance - estimate (ideal value: 0) relative to the actual variance	7 5 %	12 9 %
Correlation coefficient between XS and estimate (ideal value: 1)	0.99	0.98
Standard-deviation of the differences (ideal value: 0) relative to the mean XS value	1.9 3 %	2.2 4 %

Table 8. Some statistics on the differences between the original and synthesized images (in radiance or relative value) for Barcelona and XS1 band.

	0.001	**1**	**2**	**5**	**10**	**20**	**50**
20 m	27	28	63	92	99	100	100
40 m	23	24	56	88	99	100	100

Table 9. Probability (in per cent) for having in a pixel a relative error less or equal to the thresholds noted in the first row. The ideal value is 100 as early as the first threshold 0.001 %. The relative errors are in absolute value and in per cent. For Barcelona and XS1.

5. CONCLUSION

Many methods have been proposed for the merging of high spectral and high spatial resolutions data in order to produce multispectral images having the highest spatial resolution available within the data set. A very few propose an assessment of the quality of the resulting synthetic images. The present work proposes both a formal approach and some criteria to provide a quantitative assessment of the synthetic images. The approach is based upon simple concepts, easy to understand and easy to implement and use. Together with the visual appreciation of the synthetic images, these criteria may be used by to select a method amongst others, according to its performances for the criteria which are the most important for the application. As for the mathematicians, these criteria, somewhat extended to more complex statistical quantities, are a tool to assess the merits and drawbacks of a method under development.

This text is illustrated by the case of SPOT images. Without any problem, this work can be applied to other combinations of sensors. For example Mangolini *et al.* (1992) have assessed the performances of the ARSIS method to synthesize Landsat TM6 (thermal infrared band) at a resolution of 30 m.

6. ACKNOWLEDGMENTS

This work has been partly supported by CNES (Centre National d'Etudes Spatiales). The advice of Patrick Dorlet, Philippe Munier and Thierry Rousselin is gratefully acknowledged.

7. REFERENCES

Kong X. N., D. Vidal-Madjar, 1988, Effet de la résolution spatiale sur les propriétés statistiques des images satellites : une étude de cas. *International Journal of Remote Sensing*, **9**, 1315-1328.

Mangolini M., T. Ranchin, L. Wald, 1992, Procédé et dispositif pour augmenter la résolution spatiale d'images à partir d'autres images de meilleure résolution spatiale. French patent n° 92-13961, November 20, 1992. Methods and device for increasing the spatial resolution of images from other images of better spatial resolution. USA patent serial 08/249,882, May 26, 1994.

Mangolini M., T. Ranchin, L. Wald, 1993, Fusion d'images SPOT multispectrales (XS) et panchromatique (P), et d'images radar. *De l'optique au radar, les applications de SPOT et ERS*, pp. 199-209, Cepaduès-Editions, 111 rue Vauquelin, Toulouse, France, 574 p.

Munechika C. K., J. S. Warnick, C. Salvaggio, J. R. Schott, 1993, Resolution enhancement of multispectral image data to improve classification accuracy. *Photogrammetric Engineering & Remote Sensing,* **59**(1), 67-72.

Raffy M., 1993, Remotely-sensed quantification of covered areas and spatial resolution. *International Journal of Remote Sensing*, **14**(1), 135-159.

Ranchin T., L. Wald, M. Mangolini, 1993, Application de la transformée en ondelettes à la simulation d'images SPOT multispectrales de résolution 10 m. *In* Compte-rendus du 14ème colloque GRETSI, pp. 1387-1390.

Ranchin T., L. Wald, M. Mangolini, 1994, Efficient data fusion using wavelet transform: the case of SPOT satellite images. *In* Proceedings of the SPIE's 1993 International Symposium on Optics, Imaging and Instrumentation, vol. **2034**, pp. 171-178.

Welch R. M., M. S. Navar, S. K. Sengupta, 1989, The effect of spatial resolution upon the texture-based cloud field classifications. *Journal of Geophysical Research*, **94**(D12), 14,767-14,781.

Woodcock C. E., A. H. Strahler, 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**, 311-332.